CHAPTER 2

Performance Characteristics of Sensors and Actuators

Humans, sensing, and actuation

Beyond the natural senses and actuation in living organisms, sensing and actuation is almost exclusively a human activity whose ultimate purpose is to improve our lives and our interactions with the universe. Sensors and actuators are ubiquitous in our lives, whether we are aware of them or not. But beyond industrial sensors, those that produce many of the products we use, keep our transportation moving, and watch over our safety, there are two types of sensors and actuators that merit separate attention. The first class of devices includes those used to improve and sustain our health. From artificial limbs and organs to implantable devices, robot-assisted surgery, medical tests, and the manipulation of tissue and cells, this class of sensors and actuators is an important part of our health system and, indeed, life. They include systems such as X-ray imaging, magnetic resonance imaging (MRI), computed tomography (CT or CAT) scans, ultrasound scanning, and robotic surgery systems. Still others, of a perplexing variety, are used to test for every conceivable substance and condition in the body.

The second class of devices expands our knowledge of the universe around us and, hopefully, allows us to better understand the universe, our place in it, and ultimately to live in harmony with it. Sensing of the environment not only benefits us, but contributes to the environment itself and all organisms in it. Off the planet, sensors allow us to protect ourselves from radiation, the effects of solar flares, and maybe even to avoid catastrophic collisions with meteorites, but perhaps most of all, they satisfy our curiosity.

2.1 INTRODUCTION

Aside from the functionality of a sensor or actuator, that is, its basic function, the performance characteristics of the device or system are the most important issues the engineer is faced with. If we need to sense temperature, then of course a temperature sensor is needed. But what kind of sensor, what temperature range can it sense? How “accurate” does it need to be? Is it important that it be a linear measurement, and how critical is the repeatability of the sensor? Does it need to respond quickly or can we use a slow responding sensor? Similarly, if we use a motor to position a writing head in a printer or to machine a metal piece, what performance characteristics are important in its selection? These questions and others will be addressed here. The basic properties of sensors and actuators, that is, their performance characteristics, will be defined with a view to their interface to controllers.

The characteristics of a device start with its transfer function, that is, the relation between its input and output. This includes many other properties, such as span (or range), frequency response, accuracy, repeatability, sensitivity, linearity, reliability, and resolution, among others. Of course, not all are equally important in all sensors and actuators, and often the choice of properties will depend on the application. And it is important to have the application in mind when selecting a device, since the very best performing sensor or actuator may not be the best choice for all applications.

The properties of sensors and actuators are usually given by the manufacturer and engineers can usually rely on these data. There are instances, however, in which one might wish to use a device outside its stated range or improve on one of its properties (say, linearity), or even use it for an unintended use (e.g., use a microphone as a dynamic pressure sensor or as a vibration sensor). In these cases, the engineer will need to evaluate the characteristics or, at the very least, derive its calibration curve rather than relying on the manufacturer's calibration curve. Sometimes, too, the available data may be lacking certain information, again necessitating an evaluation. In cases such as these, the engineer needs to understand what affects these properties and what can be done to control them.

2.2 | INPUT AND OUTPUT CHARACTERISTICS

Before we can properly define input and output characteristics, it is best to first define the input and output of sensors and actuators. For a sensor, the input is the stimulus or the measured quantity (measurand). The output may be any number of quantities, including voltage, current, charge, frequency, phase, or a mechanical quantity such as displacement. For an actuator, the input is usually electric (voltage or current) and the output may be electrical or mechanical (displacement, force, a dial gauge, a light indication, a display, etc.). But one should keep in mind that the input and output may be more general. They may be mechanical or even chemical. We can describe both types of devices by a transfer function that relates input and output regardless of what these quantities are. In addition, we must take into account input and output properties such as impedance, temperature, and environmental conditions in order to provide proper operating conditions for the device.

2.2.1 Transfer Function

Also called the transfer characteristic function, the input/output characteristic function or response of a device is a relationship between the output and input of the device, usually defined by some kind of mathematical equation and a descriptive curve or graphical representation in a given range of inputs and outputs. The function may be linear or nonlinear, single valued or multivalued, and may at times be very complex. It may represent a one-dimensional relation (between a single input and a single output) or may be multidimensional (between multiple inputs and one output). In simple terms, it defines the response of a sensor or actuator to a given input or set of inputs and is one of the main parameters used in design. With the exception of linear transfer functions, it is usually difficult to describe the transfer function mathematically, although we can indicate it at least symbolically as

$$S = f(x), \tag{2.1}$$

where x is the input (stimulus in sensors or, say, current to an actuator) and S is the output. The dependence of the output S on x indicates that this function can be (and often is) nonlinear.

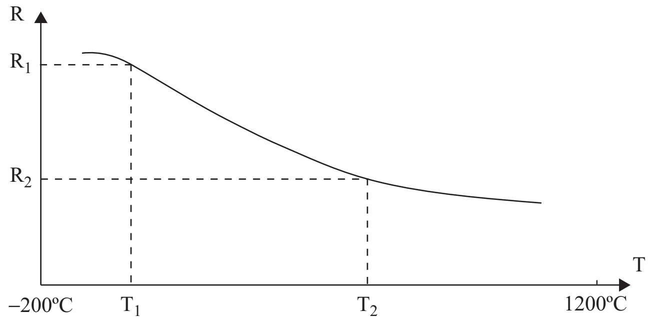

FIGURE 2.1 ■ Resistance-temperature relationship in a hypothetical temperature sensor.

Often the transfer function will be given graphically and will be limited to a range of inputs and outputs. **Figure 2.1** shows the input-output relationship for a hypothetical temperature sensor. The range between T_1 and T_2 is approximately linear and may be described by the following transfer function:

$$aT + b = R, \tag{2.2}$$

where R is the resistance of the sensor (output) and T is the temperature it senses (input) in the range $T_1 < T < T_2$.

However, the ranges below T_1 and above T_2 are nonlinear and require much more complex transfer functions, which may actually be found experimentally or may be polynomials derived by curve fitting. In many cases the sensor is restricted to operate in the linear range, in which case the graphical representation is sufficient. The curve in **Figure 2.1** contains additional data about the sensor, such as saturation, sensitivity, and range, which will be discussed in subsequent sections. It is rather rare that a transfer function is available other than in the form of a general curve or some statement as to its shape (linear, quadratic, etc.). Often it is necessary to derive it experimentally through a calibration process. There are exceptions however. Thermocouple transfer functions are available as high-order polynomials giving the transfer functions in very accurate form, as can be seen in **Example 2.1**.

EXAMPLE 2.1 Transfer function of a thermocouple

The output (voltage) of a thermocouple (temperature sensor) for a given temperature is given by a polynomial that can range from a 3rd order to a 12th order polynomial depending on the type of thermocouple. The output of a particular type of thermocouple is given by the following relation in the range 0°C–1820°C:

$$\begin{aligned} V = (&-2.4674601620 \times 10^{-1} \times T + 5.9102111169 \times 10^{-3} \times T^2 \\ &- 1.4307123430 \times 10^{-6} \times T^3 + 2.1509149750 \times 10^{-9} \times T^4 \\ &- 3.1757800720 \times 10^{-12} \times T^5 + 2.4010367459 \times 10^{-15} \times T^6 \\ &- 9.0928148159 \times 10^{-19} \times T^7 + 1.3299505137 \times 10^{-22} \times T^8) \times 10^{-3}\ \text{mV}. \end{aligned}$$

This is a rather involved transfer function (most sensors will have a much simpler response) and is nonlinear. The main purpose of the elaborate function is to provide very accurate representation over the range of the sensor (in this case 0°C–1820°C). The transfer function is shown in **Figure 2.2**.

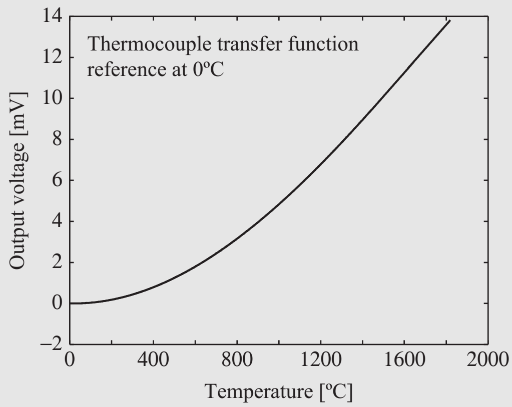

FIGURE 2.2 ■ Transfer function of the thermocouple in the range 0°C–1820°C. The output is shown in volts (V), although the polynomial gives it in millivolts (mV).

EXAMPLE 2.2 Experimental evaluation of the transfer function of a sensor

A force sensor is connected in a circuit that produces a digital output in the form of a train of pulses. The frequency of these pulses is the output of the sensor (actually of the system that includes the sensor and the circuit that converts the output to frequency—we will later call this a smart sensor). The plotted measurements are shown in **Figure 2.3**. The input force range is 0–7.5 N, for which the output produces a frequency between 25.98 kHz and 39.35 kHz. The range between 1 N and about 5 N may be useful as a "linear" range provided that the error incurred by doing so is acceptable. Below 1 N and above 6 N the output is not usable because of reduced response (saturation).

Note: With appropriate circuitry the response of a sensor (or actuator) may be linearized to a large degree if that is deemed useful. Also, it should be noted that the nonlinearity of the curve as well as its saturation may be due to the sensor itself, the electronics in the circuit, or both.

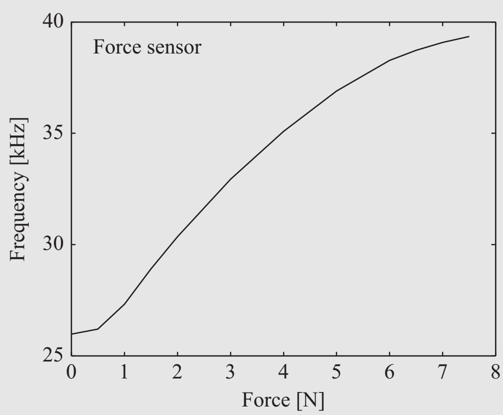

FIGURE 2.3 ■ Experimental evaluation of the transfer function of a force sensor.

Another type of input-output characteristic function, sometimes called a transfer function, is the frequency response of a device. It may be called a transfer function because it gives the output response to input over a frequency range. It is sometimes called a frequency transfer function. We will simply call it a frequency response, and it will be discussed separately in **Section 2.2.7**.

The inputs and outputs of sensors or actuators may be further characterized by the type of signal they provide or require. The output signal of a sensor may be voltage or current or it may be frequency, phase, or any other measurable quantity. The output of actuators is usually mechanical, manifest in motion or force, but it may be in other forms such as light or electromagnetic waves (antennas are actuators when they transmit or sensors when they receive), or it may be chemical. Within these characteristics, some devices operate at very low or very high levels. For example, the output of a thermocouple is typically 10–50 μV/°C, whereas a piezoelectric sensor can produce 300 V or more in response to motion. A magnetic actuator may require, say, 20 A at 12 V, whereas an electrostatic actuator may operate at 500 V (or higher) at a very low current.

2.2.2 Impedance and Impedance Matching

All devices have an internal impedance that may be real or complex. Although we can view both sensors and actuators as two-port devices, we will only be concerned here with the output impedance of sensors and the input impedance of actuators, since these are the properties necessary for interfacing and are readily measured. The range of impedances of devices is large and the importance of impedance matching cannot be overstated; more often than not, failure to properly match a device means failure of the sensing strategy, or in the case of actuators, it may mean physical destruction of the actuator, its drivers, or worse.

The input impedance of a device is defined as "the ratio of the rated voltage and the resulting current through the input port of the device with the output port open" (no load). The output impedance is defined as "the ratio of the rated output voltage and the short circuit current of the port" (i.e., the current when the output is shorted).

The reason these properties matter is that they affect the operation of the device. To understand this, consider first the output impedance of a hypothetical strain gauge (a strain sensor) that has an output resistance of 500 Ω at no strain and 750 Ω at a higher strain. The strain gauge is an active sensor so we must connect it to a source as in **Figure 2.4a**. As the strain increases, the sensor's resistance increases. The strain is measured by measuring this change in resistance in terms of the change in voltage on the sensor. At no strain, this voltage is 2.5 V (corresponding to 500 Ω) and at the measured strain it is 3 V (corresponding to 750 Ω). Suppose now we connect this sensor to a processor, which has an input impedance of 500 Ω as well. As soon as the connection is made, the voltage across the sensor goes down to 1.666 V at no strain and rises to 1.875 V at the measured strain (see **Figure 2.4b** and **2.4c**). Two things should be noted here. First is the reduction in the output of the sensor from 2.5 V to 1.666 V. We refer to this as **loading** of the sensor by the input impedance of the processor. Second, and more importantly, whereas the output at no load rose by 0.5 V, when connected it rose by only 0.209 V. This change may be viewed as a reduction in sensitivity (see **Section 2.2.5**), and unless special measures are taken, it may result in an erroneous measurement of strain. In this case, the obvious solution is for the input impedance of the processor to be

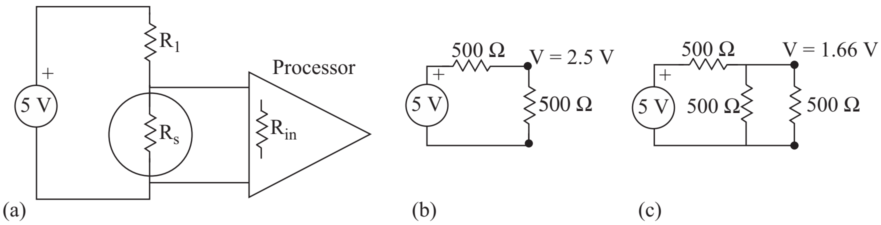

FIGURE 2.4 ■ A strain sensor connected to a processor. (a) The sensor and connections. (b) Equivalent circuit of a sensor alone. (c) Equivalent circuit of a sensor and processor.

as high as possible (ideally infinite) or an impedance matching circuit can be connected between the sensor and the processor. This matching circuit must have very high input impedance and low output impedance. We shall see later that such circuits are available and commonly used.

On the other hand, if the output of a sensor is current, it must be connected to as small an external impedance as possible to avoid changes in the sensor's current, or again a matching circuit with low input and high output impedance must be used. The same exact considerations apply to actuators.

EXAMPLE 2.3 Force sensor

A force sensor is used in an electronic scale to weigh items ranging from 1 gf to 1000 gf. The sensor's resistance changes linearly from 1 MΩ to 1 kΩ as the force changes from 1 gf to 1000 gf (9.80665 mN to 9.80665 N). To measure the resistance, the sensor is connected to a constant current source of 10 μA and the voltage across the sensor is used as the measured quantity. The voltage is measured with a voltmeter with internal impedance of 10 MΩ. The configuration is shown in **Figure 2.5a**. Calculate the error produced by the connection of the voltmeter. What will the actual readings be for the range of forces given?

Solution: The voltage on the sensor changes from 10 V for a mass of 1 g to 0.01 V for a mass of 1000 g before the voltmeter is connected. When the voltmeter is connected, the net resistance is lower and the voltage across the sensor is lower as well. The voltmeter adds its impedance in parallel with the sensor's as shown in **Figure 2.5b**. The net resistances at the ends of the range are at 1 g,

$$R(1\text{ g}) = \frac{R_s R_v}{R_s + R_v} = \frac{10^6 \times 10^7}{10^6 + 10^7} = \frac{10}{11} \times 10^6 = 0.909090 \times 10^6\ \Omega,$$

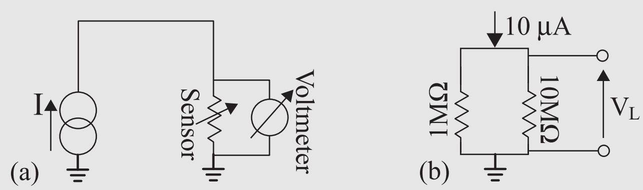

FIGURE 2.5 ■ Loading of a sensor by the measuring instrument. (a) The circuit. (b) The equivalent circuit.

and at 1000 g,

$$R(1000\ \text{g}) = \frac{R_s R_v}{R_s + R_v} = \frac{10^3 \times 10^7}{10^3 + 10^7} = \frac{1}{1.001} \times 10^3 = 0.99900 \times 10^3\ \Omega.$$

The measured voltages are

$$V(1\ \text{g}) = I_s R(1\ \text{g}) = 10 \times 10^{-6} \times 0.909090 \times 10^6 = 9.09\ \text{V}$$

and

$$V(1000\ \text{g}) = I_s R(1000\ \text{g}) = 10 \times 10^{-6} \times 0.99900 \times 10^3 = 0.00999\ \text{V}.$$

Clearly the error at 1 g is 9.1%, whereas at 1000 g the error is 0.1%. The actual readings are 0.909 g and 999 g. The error at 1000 g is small and acceptable but at 1 g it may be too high. If higher accuracy is needed, the impedance of the voltmeter must be higher still. This can again be done through electronic circuits, as we shall see in **Chapter 11**.

In some cases, especially in actuators, the output is power rather than voltage or current. In such cases we are usually interested in transferring maximum power from the processor to the actuated medium (say, e.g., from an amplifier to air through a loudspeaker). Maximum power transfer is achieved through conjugate matching, which means simply that given an output impedance of a processor, $R + jX$, the input impedance of the actuator must be equal to $R - jX$. In the case of resistances (real impedances), conjugate matching means that the output resistance of the processor and the input resistance of the actuator must be the same. A very simple example of this may be found in audio amplifiers; an amplifier will transfer maximum power to an 8 Ω speaker if the amplifier's output equals 8 Ω (however, see **Example 2.5**). Although not common, there are sensors that also operate in this mode, in which case conjugate matching applies to them as well.

EXAMPLE 2.4 Impedance matching in actuators

A voice coil actuator (an actuator that operates on the principle of the loudspeaker; we shall discuss these in **Chapter 5**) is pulse driven by an amplifier. The amplifier provides an amplitude $V_s = 12$ V. The internal impedance of the amplifier is $R_s = 4\ \Omega$.

a. Calculate the power transferred to an impedance-matched actuator.
b. Show that the power transmitted to an actuator with lower or higher impedance is lower than that for the matched actuator in (a).
c. What is the power supplied to a 4 Ω actuator if the internal impedance of the amplifier is 0.5 Ω and supplies the same voltage (12 V)?

Solution: Because the actuator is pulse driven, the power is considered instantaneous, but since the voltage is constant through the duration of the pulse, we will calculate the power as if it were a DC source (i.e., the power during the ON portion of the pulse).

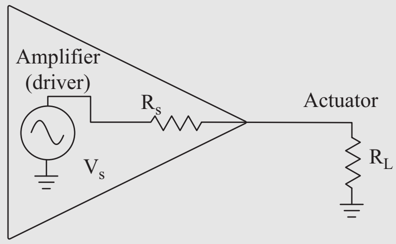

FIGURE 2.6 ■ The concept of matching the load impedance to the internal impedance of a processor.

a. The equivalent circuit for the matched condition is shown in **Figure 2.6**. The actuator's resistance is $R_L = 4\ \Omega$ and the power supplied to the actuator is

$$P_L = \frac{V_L^2}{R_L} = \left(\frac{V_s}{R_s + R_L}R_L\right)^2 \frac{1}{R_L} = \left(\frac{12}{4+4}4\right)^2 \frac{1}{4} = 9\ \text{W}.$$

Note that this is exactly half the power supplied by the source, the other half being dissipated on the internal resistance of the source in the form of heat.

b. With a lower or higher actuator impedance we have, for $R_L = 2\ \Omega$ for example,

$$P_L = \frac{V_L^2}{R_L} = \left(\frac{V_s}{R_s + R_L}R_L\right)^2 \frac{1}{R_L} = \left(\frac{12}{4+2}2\right)^2 \frac{1}{2} = 8\ \text{W}.$$

Similarly, for a higher actuator impedance, say for $R_L = 6\ \Omega$,

$$P_L = \frac{{V_L}^2}{R_L} = \left(\frac{V_s}{R_s + R_L}R_L\right)^2 \frac{1}{R_L} = \left(\frac{12}{4+6}6\right)^2 \frac{1}{6} = 8.64\ \text{W}.$$

Clearly, maximum power is transferred for matched impedances.

Note: The exact condition for matching is that of $Z_L = Z_s^*$, that is, if Z_s is complex, $Z_s = R_s + jX_s$, then a matched load has impedance $Z_L = R_L - jX_s$. In the case shown here, $X_s = 0$; hence the matching condition is $R_L = R_s$.

c.

$$P_L = \frac{V_L^2}{R_L} = \left(\frac{V_s}{R_s + R_L}R_L\right)^2 \frac{1}{R_L} = \left(\frac{12}{0.5+4}4\right)^2 \frac{1}{4} = 28.44\ \text{W}.$$

Note that as the internal impedance of the amplifier approaches zero, the power delivered to the load approaches 36 W and the power dissipated on the internal impedance of the amplifier approaches zero. This may seem to contradict the maximum power transfer condition. Note, however, that if the load were equal to the internal impedance (0.5 W in this case), the power delivered to the load would have been 288 W.

There are also sensors and actuators that operate at high frequencies. Impedance matching to these devices is a much more complex issue and will be discussed briefly in **Chapter 9**. It suffices to say here that the general requirement is that the connection of a sensor or actuator produces no reflection of voltage or current. In such cases the impedance of the sensor or actuator must equal the input impedance of the processor. This requirement does not guarantee maximum power transfer, only nonreflection.

2.2.3 Range, Span, Input and Output Full Scale, Resolution, and Dynamic Range

The ***range*** of a sensor refers to the lower and upper limit operating values of the stimulus, that is, the minimum and maximum input for which a valid output is obtained. Typically we would say that the range of the sensor is between the minimum and maximum values. For example, a temperature sensor may operate between $-45°C$ and $+110°C$. These are the range points.

The ***span*** of a sensor is the arithmetic difference between the highest and lowest values of the stimulus that can be sensed within acceptable errors (i.e., the difference between the range values). This may also be called the ***input full scale*** (IFS) of the sensor. The ***output full scale*** (OFS) is the difference between the upper and lower ranges of the output of the sensor corresponding to the span of the sensor. For example, a sensor measures temperature between $-30°C$ and $+80°C$ and produces an output between 2.5 V and 1.2 V. The span (IFS) is $80°C - (-30°C) = 110°C$ and the OFS is $2.5 V - 1.2 V = 1.3 V$. The range of the sensor is between $-30°C$ and $+80°C$. The IFS and OFS apply equally to actuators.

The range and span of a sensor or actuator express essentially the same information in slightly different ways and therefore they are often used interchangeably.

The **resolution** of a sensor is the minimum increment in stimulus to which it can respond. It is the magnitude of the input change that results in the smallest discernible output. For example, a sensor may be said to have a resolution of 0.01°C, meaning that an increment in temperature of 0.01°C produces a readily measurable output. In common usage this is sometimes, erroneously, referred to as sensitivity. Resolution and sensitivity are two very distinct properties (**sensitivity** is the ratio of change in output to the change in input and will be discussed separately). Resolution of analog devices is said to be infinitesimal, that is, their response is continuous and hence the resolution depends on our ability to discern the change. Often the resolution is defined by the noise level, since for a signal to be discernible it must be larger than the noise level. The resolution of a sensor or actuator is often defined by the instrument or processor used to measure the output. For example, a sensor producing, say, an output of 0–10 V for a temperature range of 0°C–100°C must be connected to an instrument to monitor that voltage. If this instrument is analog (an analog voltmeter, for example), the resolution may be, say, 0.01 V (1000 graduations on the voltmeter's scale) or 0.001 V (10,000 graduations). Even then, when reading the voltmeter, one may interpolate between the graduations, extending the resolution. If, however, the voltmeter is digital and it has an increment of 0.01 V, then this is the resolution of the instrument, and by extension, that of the sensing system made of the sensor and the instrument.

Resolution may be specified in the units of the stimulus (e.g., 0.5°C for a temperature sensor, 1 mT for a magnetic field sensor, 0.1 mm for a proximity sensor, etc.) or may be specified as a percentage of the span (0.1%, for example).

The resolution of an actuator is the minimum increment in its output that it can provide. For example, a DC motor is capable of infinitesimal resolution, whereas a stepper motor may have 200 steps/revolution for a resolution is 1.8°.

In digital systems, resolution may be specified in bits (such as N-bit resolution) or in some other means of expressing the idea of resolution. In an analog-to-digital (A/D) converter, the resolution means the number of discrete steps the converter can convert. For example, a 12-bit resolution means the device can resolve $2^{12} = 4096$ steps. If the converter digitizes a 5 V input, each step is $5/4096 = 1.22 \times 10^{-3}$ V. On the analog side, the resolution may be described as 1.22 mV, but on the digital side it is described as 12-bit resolution. In digital cameras and in display monitors the resolution is typically given as the total number of pixels. Thus a digital camera may be said to have a resolution of a number (x) of megapixels.

EXAMPLE 2.5 Resolution of a system

Suppose a signal is digitized and measured with a two-digit digital voltmeter capable of measuring up to 1 V (after proper amplification). The possible measurement is between 0 and 0.99 V with a resolution of 0.01 V or 1%. In this case the resolution is limited by the voltmeter (the actuator in this system), whereas the signal is continuous, and given a "better" voltmeter, we may well be able to resolve it further. We shall see in **Chapter 11** that A/D converters are capable of much higher resolution than the one shown here.

EXAMPLE 2.6 Resolution of anolog and digital sensors

A digital pressure sensor has a range between 100 kPa (approximately 1 atm) and 1 MPa (approximately 10 atm). The sensor is an analog sensor and produces an output voltage that varies between 1 V at the lowest range and 1.8 V at the highest range. The digital output display shows the pressure directly using a 3½-digit panel meter (a 3½-digit panel meter has three digits that can display 0 to 9 and one digit that can display 0 and 1). What is the resolution of the sensor itself (analog) and what is the resolution of the digital sensor, assuming that the display is autoranging, that is, it is capable of placing the digital point automatically?

Solution: The analog sensor has infinitesimal resolution and its output changes continuously between 1 V and 1.8 V. That is, for a change in pressure of 9×10^6 Pa, the voltage changes by 0.8 V or 88.89 nV/Pa. Of course, the pascal is a very small unit, so we might say as well that the output changes by 88.89 μV/kPa. That is, if we were to assume that an output of 100 nV can be read reliably, the resolution of the sensor is 100/88.89 = 1.125 Pa. On the other hand, if the output can only read, say, 10 μV, the resolution is 10,000/88.89 = 112.5 Pa. The practical resolution is limited only by our ability to measure voltage and by any noise that may exist.

The digital panel meter will display from 0.100 to 10.00 MPa. In the range 0.100 to 9.999 MPa, the resolution is clearly 0.001 MPa, or 1 kPa. At 10.00 MPa the resolution decreases to 0.01 MPa, or 10 kPa. Note that the limitation is imposed by the display itself.

The **dynamic range of a device** (sensor or actuator) is the ratio of the span of the device and the minimum discernible quantity the device is capable of (resolution). Typically the lowest discernible value is taken as the noise floor, that is, the level at which the signal is "lost" in the noise. The use of dynamic range is particularly useful in devices with large spans and for that reason is usually expressed in decibels. Since the ratio represents either powerlike (power, power density) or voltagelike quantities (voltage, current, force, fields, etc.), the dynamic range is written as follows:

For voltage-like quantities:

$$\text{Dynamic range} = 20 \log|\text{span}/\text{lower measurable quantity}|. \tag{2.3}$$

For power-like quantities:

$$\text{Dynamic range} = 10 \log|\text{span}/\text{lower measurable quantity}|. \tag{2.4}$$

Suppose we look at a 4-digit digital voltmeter capable of measuring between 0 and 20 V. The total span is 19.99 V and the resolution (smallest increment) is 0.01 V. The dynamic range is thus

$$\text{Dynamic range} = 20 \log(19.99/0.01) = 20 \times 3.3 = 66 \text{ dB}.$$

On the other hand, a 4-digit digital wattmeter measuring up to 20 W in increments of 0.01 W will have a dynamic range of

$$\text{Dynamic range} = 10 \log(19.99/0.01) = 10 \times 3.3 = 33 \text{ dB}.$$

Span, IFS, and OFS are usually measured in the respective quantity at the input and output of the device (pressure and voltage in the example above), but in some cases, where the dynamic range is very large, these may also be given in decibels. The same consideration may be applied to actuators, but whereas dynamic range is often used for sensors, the inputs and outputs of actuators are usually defined by the range, especially when motion is involved.

EXAMPLE 2.7 **Dynamic range of a temperature sensor**

A silicon temperature sensor has a range between 0°C and 90°C. The accuracy is defined in the data sheet as ±0.5°C. Calculate the dynamic range of the sensor.

The resolution is not given, so we will take the accuracy as the minimum measurable quantity. In general, these need not be the same. Since the minimum resolution is 0.5°, the dynamic range is

$$\textit{Dynamic range} = 20\log_{10}\left(\frac{90}{0.5}\right) = 45.1 \text{ dB}.$$

EXAMPLE 2.8 **Dynamic range of a loudspeaker**

A loudspeaker is rated at 6 W and requires a minimum power of 0.001 W to overcome internal friction. What is its dynamic range? Clearly, any change smaller than 1 mW will not change the position of the speaker's cone and hence no change in output will be produced. Thus 1 mW is the resolution of the speaker and the dynamic range of the speaker is

$$\textit{Dynamic range} = 10\log_{10}\left(\frac{6}{0.001}\right) = 37.78 \text{ dB}.$$

In digital sensors and actuators, the signal levels change in increments of bits. In general, one can define a dynamic range based on the digital representation or on the equivalent analog signal. In an N-digit device, the ratio between the highest and lowest representable level is $2^N/1 = 2^N$. Therefore the dynamic range may be written as

$$\textit{Dynamic range} = 20\log_{10}\left(2^N\right) = 20N\log_{10}(2) = 6.0206N \quad [\text{dB}]. \tag{2.5}$$

EXAMPLE 2.9 **Dynamic range of an A/D converter**

A 16-bit analog to digital [A/D] converter (to be discussed in **Chapter 11**) is used to convert an analog music recording into digital format so it can be stored digitally and played back (by converting it back to analog form). The amplitude varies between -6 V and $+6$ V.

a. Calculate the smallest signal increment that can be used.
b. Calculate the dynamic range of the A/D conversion.

Solution:
a. A 16-bit A/D converter can represent $2^{16} = 65{,}536$ levels of the signal. The smallest signal increment in the case discussed here is

$$\Delta V = \frac{12}{2^{16}} = 1.831 \times 10^{-4}\ \text{V}.$$

That is, the signal is represented in increments of 0.1831 mV.

b. The dynamic range of the A/D converter is

$$\textit{Dynamic range} = 6.0206N = 96.33\ \text{dB}.$$

2.2.4 Accuracy, Errors, and Repeatability

The errors involved in sensing and actuating define the accuracy of the device. These may stem from various sources, but they all represent deviations of the device output from the ideal. Inaccuracies in the output (i.e., in the transfer function) stem from a variety of sources, including materials, construction tolerances, aging, operational errors, calibration errors, matching (impedance) or loading errors, and many others. The definition of error is rather simple and can be represented as the difference between the measured and actual value. In practice, it may be represented in a number of ways.

1. The most obvious is as a difference: $e = |V - V_0|$, where V_0 represents the actual (correct) value and V is that measured by the device. Often the error is given as $\pm$e.
2. A second way is to represent this as a percentage of IFS (span), $e = (\Delta t/(t_{\max} - t_{\min})) \times 100$, where $t_{\max}$ and $t_{\min}$ are the maximum and minimum values at which the device is designed to operate (range values).
3. A third method is to specify the error in terms of the output signal expected rather than the stimulus. Again, it may simply be the difference between values or it may be represented as a percentage of OFS.

EXAMPLE 2.10 **Errors in sensing**

A thermistor is used to measure temperatures between $-30°C$ and $+80°C$ and produces an output voltage between 2.8 V and 1.5 V. The ideal transfer function is shown in **Figure 2.7** (solid line). Because of errors, the accuracy in sensing is $\pm 0.5°C$. The errors may be specified as follows:

a. In terms of the input as $\pm 0.5°C$.

b. As a percentage of the input range:

$$e = [0.5/(80 + 30)] \times 100 = 0.454\%.$$

c. In terms of the output range. This may be taken off the curve as the difference shown or may be calculated by first evaluating the transfer function and its maximum and minimum limits, as shown in **Figure 2.7**. This gives an error of ± 0.059 V. It may also be given in percentage of OFS as $\pm[0.059/(2.8 - 1.5)] \times 100 = 4.54\%$. Note that these are not the same and the appearance of accuracy will depend on how it is expressed. In most cases, the error given in terms of the measurand or percent of IFS is the best measure of the sensor's accuracy.

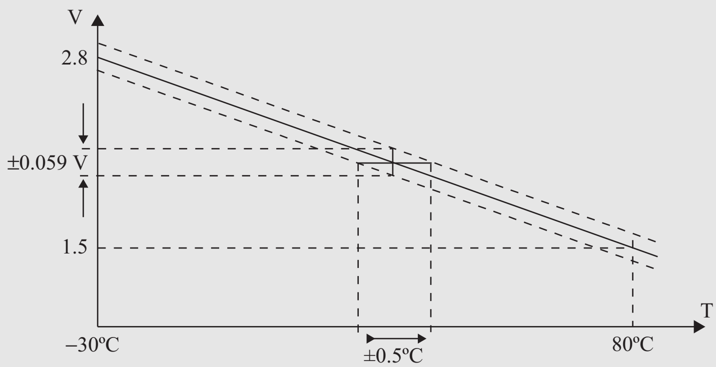

FIGURE 2.7 ■ Transfer function and error limits.

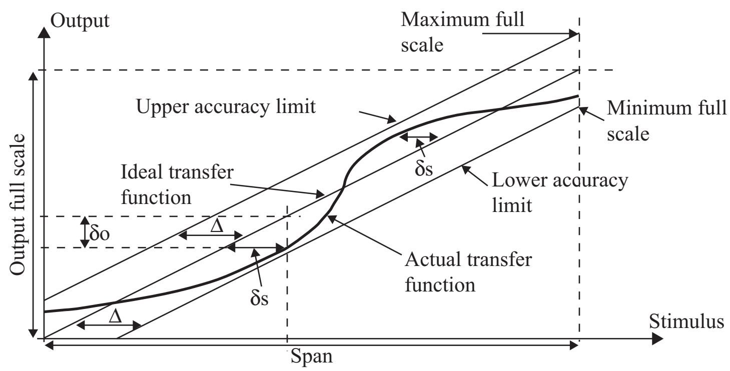

FIGURE 2.8 ■ Error and accuracy limits in nonlinear transfer functions.

In general, the transfer function is nonlinear and the error varies throughout the range of the sensor, as illustrated in **Figure 2.8**. When this happens, either the maximum error or an average error are taken as representative of the error of the device. **Figure 2.8** shows the error limits or accuracy limits as parallel curves that limit the actual transfer

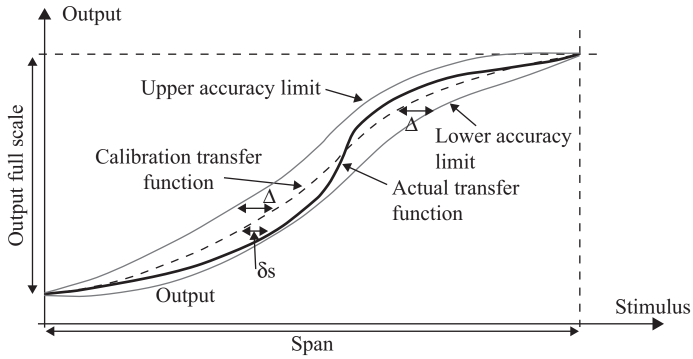

FIGURE 2.9 ■ Error in nonlinear transfer functions may be taken around the calibration transfer function rather than the ideal transfer function.

function. The accuracy limits are parallel to the ideal transfer function and do not have to be straight lines (see **Figure 2.9**). In some cases, the sensor may be calibrated during production or during installation. In such cases, the error may be taken around the calibration curve rather than around the ideal transfer function, as shown in **Figure 2.9**. In effect, rather than using the ideal transfer function, the actual transfer function (calibration curve) is used. This has the effect of being more realistic for the specific device, but does not apply to other devices of the same type—each must be calibrated separately.

So far errors have been assumed to be static. That is, they do not depend on time. However, errors can also be dynamic or time dependent, but the calculation and meaning of errors is the same as for static errors.

Some errors are random, whereas others are constant or **systemic**. If various samples of a device exhibit different errors in a particular parameter, or if a particular device exhibits different error values each time it is operated, these errors are said to be random. If the errors are constant, they are said to be systemic. A device may have both systemic and random errors.

EXAMPLE 2.11 Nonlinearity errors

In a capacitive accelerometer the distance between two plates is related to the force due to acceleration, $F = ma$, where m is the mass of the moving plate and a is the acceleration (see **Figure 2.10a**). A spring generates the restoring force. The relation between the force and the distance between the plates (hence capacitance) is determined experimentally and is given in the table below. Establish the maximum error due to the nonlinearity of the response.

d [mm]	0.52	0.5	0.48	0.46	0.44	0.42	0.4	0.38	0.36	0.34	0.32	0.3	0.28	0.26
F [μN]	0	6	9	13	17	21	25	28	31	35	39	43	46	49

d [mm]	0.24	0.22	0.2	0.18	0.16	0.14	0.12	0.1	0.08	0.06	0.04	0.02	0.012	0
F [μN]	52	55	58	61	64	67	70	73	76	79	82	84	85	86

Solution: The transfer function and the upper and lower accuracy limits are plotted in **Figure 2.10b**. The maximum error is 8.0 μN and occurs at the very beginning and end of the curve (i.e., at $d = 0$ and at $d = 0.520$ mm. If we disregard these as the extremes, at which the sensor is likely to be less accurate, the maximum error is 7.6 μN and occurs at $d = 0.22$ mm.

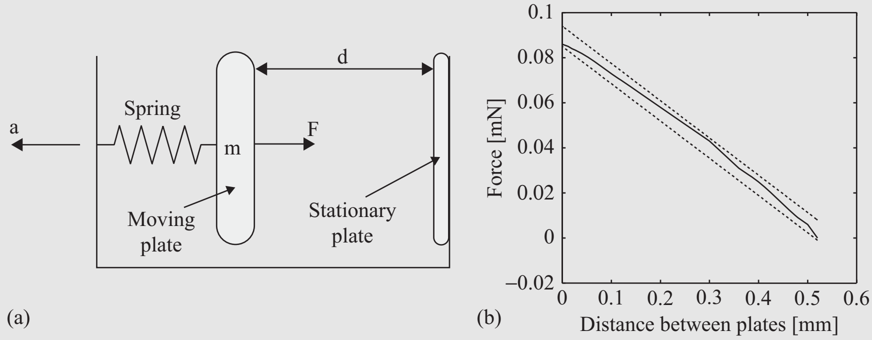

FIGURE 2.10 ■ (a) Capacitive accelerometer. (b) Transfer function of the accelerometer in (a) showing the upper and lower accuracy limits.

The error in sensed distance is 0.04 mm. These errors are derived from the accuracy limits shown since the ideal transfer function is linear ($F = ma = kx$, where k is the spring constant and x is the displacement). The errors may equally be given in percentage of full scale. The error in displacement is $(0.04/0.52) \times 100 = 7.69\%$, whereas the error in force is $(7.6/86) \times 100 = 8.84\%$.

Note: Typically the error would be taken from the ideal transfer function to the accuracy limit. Here, however, we look at the error due to nonlinearity, hence the error is taken between the two accuracy limits that bound the transfer function. In effect, the error is taken as if the ideal transfer function were the lower accuracy limit.

Repeatability, sometimes called reproducibility, of sensors and actuators is an important design characteristic and simply indicates the failure of the sensor or actuator to represent the same value (i.e., stimulus for sensors or output for actuators) under identical conditions when measured at different times. It is usually associated with calibration and is viewed as an error. It is given as the maximum difference between two readings (either two calibration readings or two measurement readings) taken at different times under identical input conditions. Usually the error will be given as a percentage of IFS.

2.2.5 Sensitivity and Sensitivity Analysis

The sensitivity of a sensor or actuator is defined as the change in output for a given change in input, usually a unit change in input. Clearly, sensitivity represents the slope of the transfer function. We can write

$$s = \frac{dS}{dx} = \frac{d}{dx}(f(x)). \tag{2.6}$$

For the linear transfer function in **Equation (2.2)**, where the output is resistance (R) and the input is temperature (T), we have

$$s = \frac{dR}{dT} = \frac{d}{dT}(aT + b) = a \quad [\Omega/^\circ\mathrm{C}]. \tag{2.7}$$

Note in particular the units: in this case, since the output is in ohms, and the stimulus is in degrees Celsius, the sensitivity is given in ohms per degree Celsius ($\Omega/^\circ$C).

Sensitivity may be constant throughout the span (linear transfer function), it may be different in different regions, or it may be different at every point in the span (such as in **Figure 2.8**).

Usually sensitivity is associated with sensors. However, as long as a transfer function can be defined for an actuator, the same ideas can be extended to actuators. Therefore it is quite appropriate to define a sensitivity of, say, a speaker as dP/dI, where P is the pressure (output) the speaker generates per unit current I (input) into the speaker, or the sensitivity of a linear positioner as dl/dV, where l is linear distance and V is the voltage (input) to the positioner.

Sensitivity analysis is usually a difficult task, especially when, in addition to the stimulus, there is noise. In addition, the sensitivity of a sensor is often a combined function of sensitivities of various components of the sensor, including the transduction section or sections (if multiple transduction steps are involved). Further, the sensor may be rather complex with multiple transduction steps, each one with its own sensitivity, sources of noise, and other parameters that come into play, such as nonlinearities, accuracy, and others, some of which may be known, but many of which may not be known or may only be known approximately. Nevertheless, sensitivity analysis is an important step in the design process, especially when complex sensors are used, since, in addition to providing information on the output range of signals one can expect, it also provides information on noise and errors. Sensitivity analysis may also provide clues as to how the effects of noise and errors can be minimized by proper choice of sensors, their connections, and other steps that can be taken to improve performance (amplifiers, feedback, etc.).

EXAMPLE 2.12 Sensitivity of a thermocouple

Consider the thermocouple discussed in **Example 2.1**. The transfer function is given as

$$\begin{aligned} V = (&-2.4674601620 \times 10^{-1} \times T + 5.9102111169 \times 10^{-3} \times T^2 - 1.4307123430 \\ &\times 10^{-6} \times T^3 + 2.1509149750 \times 10^{-9} \times T^4 - 3.1757800720 \times 10^{-12} \times T^5 \\ &+ 2.4010367459 \times 10^{-15} \times T^6 - 9.0928148159 \times 10^{-19} \times T^7 + 1.3299505137 \\ &\times 10^{-22} \times T^8) \times 10^{-3}\text{mV}. \end{aligned}$$

The sensitivity of the sensor can be found by direct differentiation:

$$\begin{aligned} s = \frac{dV}{dT} = (&-2.4674601620 \times 10^{-1} + 1.182042223 \times 10^{-2} \times T - 4.292137029 \\ &\times 10^{-6} \times T^2 + 8.6036599 \times 10^{-9} \times T^3 - 1.587890036 \times 10^{-11} \times T^4 \\ &+ 1.406220476 \times 10^{-14} \times T^5 - 6.3649703711 \times 10^{-18} \times T^6 \\ &+ 1.06396041 \times 10^{-21} \times T^7) \times 10^{-3}\text{mV}/^\circ\text{C}. \end{aligned}$$

However, this may not be as useful as it looks. Indeed, one can obtain the sensitivity at any temperature by simply substituting the temperature in this relation. But suppose we need to use the sensor to measure temperature between 0°C and 150°C. It may be more useful to obtain a single "average" value for sensitivity by first passing a linear best fit between the points on the transfer function (see **Appendix A**). Once that is done, the sensitivity is the slope of the transfer function. The steps are as follows:

1. We first obtain the output voltage using the transfer function at a number of points (the more points, the better).
2. **Equation (A.15)** is then used to obtain a linear best fit of the form $V = aT + b$.
3. The sensitivity of this linearized transfer function is a.

By generating the values of V between $T = 0°C$ and $T = 150°C$ and using **Equation (A.15)** or **Equation (2.19)** the linear transfer function is

$$V = aT + b = 6.15540978 \times 10^{-4}T - 0.02122939 \text{ mV}.$$

This is plotted together with the exact curve (eighth-order polynomial above) in **Figure 2.11**. The sensitivity now becomes

$$s = \frac{dV}{dt} = a = 6.15540978 \times 10^{-4} \text{ mV}/°\text{C}.$$

This is a low sensitivity of only 6.1554 μV/°C, but is not out of line in thermocouples.

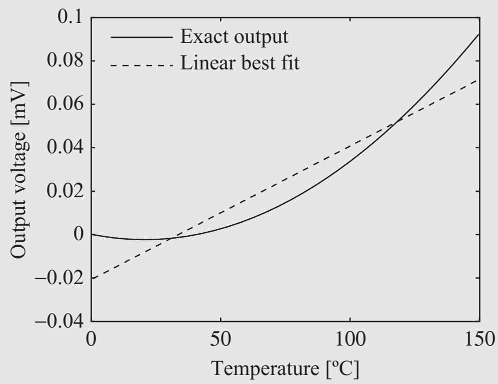

FIGURE 2.11 ■ Transfer function of a thermocouple and the linear best fit approximation.

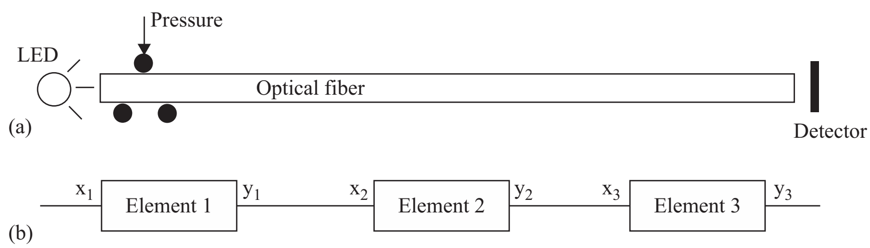

FIGURE 2.12 ■ A sensing system made of a source, optical fiber pressure sensor, and processor. (a) The sensor. (b) Equivalent configuration showing the transducing elements.

To understand some of the issues involved in sensitivity analysis, consider a sensor with three conversion steps in series. An example of a sensor of this type is the optical fiber pressure sensor shown schematically in **Figure 2.12a**. The operation is as follows: The optical fiber transmits light generated by a laser or LED to a detector and the phase of this signal is calibrated to read pressure. When pressure is applied on the optical fiber, it is tensioned (i.e., its length changes slightly). This means the light travels a longer distance in the fiber and its phase at the detector will be larger. This is a complex sensor that includes three transduction steps. First, an electric signal is converted into light and is coupled into the fiber. Then pressure is converted to displacement and, in the detector, light is converted into an electric signal. Each one of these transduction steps has its own errors, its own transfer function, and its own sensitivity.

The three transducers are connected in series and their errors are additive. The sensitivity of each element is

$$s_1 = \frac{dy_1}{dx_1}, \quad s_2 = \frac{dy_2}{dx_2}, \quad s_3 = \frac{dy_3}{dx_3}, \tag{2.8}$$

where y_i is the output of transducer i and x_i is its input. Suppose first that there are no errors in the system. Then we can write

$$S = s_1 s_2 s_3 = \frac{dy_1}{dx_1}\frac{dy_2}{dx_2}\frac{dy_3}{dx_3}. \tag{2.9}$$

But clearly $x_2 = y_1$ (the output of transducer 1 is the input to transducer 2) and $x_3 = y_2$. With these we get

$$S = s_1 s_2 s_3 = \frac{dy_3}{dx_1}. \tag{2.10}$$

This is both simple and logical. The internal conversion steps are not seen in **Equation (2.10)**, meaning that we only need to take into account the sensor as a unit with input and output.

If errors or noise are present, and assuming each transducer element has different errors, we can write the output of transducer element 1 as $y_1 = y_1^0 + \Delta y_1$, where y_1^0 is the output without errors. Assuming that we know the sensitivities of the elements, we can write the output of element 2 as

$$y_2 = s_2\left(y_1^0 + \Delta y_1\right) + \Delta y_2 = y_2^0 + s_2 \Delta y_1 + \Delta y_2, \tag{2.11}$$

where $y_2^0 = s_1 y_1^0$ is the output of element 2 without errors and Δy_2 is the errors introduced by element 2. Now this becomes the input to element 3 and we can write

$$y_3 = s_3\left(y_2^0 + s_2 \Delta y_1 + \Delta y_2\right) + \Delta y_3 = y_3^0 + s_2 s_3 \Delta y_1 + s_3 \Delta y_2 + \Delta y_3. \tag{2.12}$$

The last three terms are errors, and these are summed as they propagate through the series elements. Clearly, then, the errors in the output depend on the intermediate transduction steps.

Consider now a differential sensor designed to measure pressure difference between two locations in a system, as shown in **Figure 2.13a**. **Figure 2.13b** shows the transfer functions, inputs, and outputs. Assuming first that there are no errors and that each transducer has a different transfer function, the sensitivity of each sensor is

$$s_1 = \frac{dy_1}{dx_1}, \quad s_2 = \frac{dy_2}{dx_2}. \tag{2.13}$$

The outputs of the two sensors are $y_1 = s_1 x_1$ and $y_2 = s_2 x_2$ and the overall output is

$$y = y_1 - y_2 = s_1 x_1 - s_2 x_2. \tag{2.14}$$

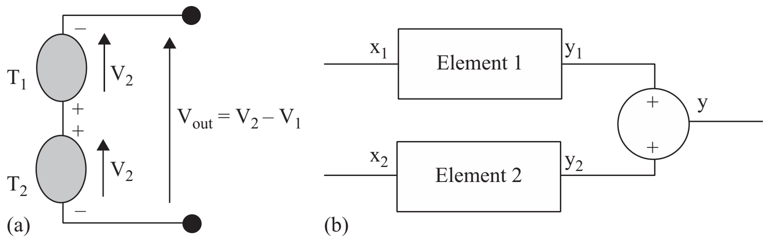

FIGURE 2.13 ■ A differential sensor. (a) The two sensors measure temperatures at different locations connected in opposition. (b) Equivalent configuration showing the transducer elements.

If the two sensors are identical, so that $s_1 = s_2 = s$, we get

$$y = s(x_1 - x_2), \tag{2.15}$$

and the sensitivity of the differential sensor can be calculated as

$$s = \frac{d(y_1 - y_2)}{d(x_1 - x_2)}. \tag{2.16}$$

If the two sensors are identical (assumed), the errors produced by them will be identical (or nearly so). Hence, in taking the difference between the two outputs, the errors cancel out and the differential sensor is virtually error free. Noise, which is one source of errors, cancels as well, as long as it is common to both sensors (common-mode noise). In practice, total cancellation does not occur because of mismatching between the two sensors and other effects, such as the fact that they may be installed at different locations and hence experience different conditions.

EXAMPLE 2.13 Sensitivity to noise

The response of a pressure sensor is determined experimentally and given in the table below. A noise in the form of pressure variations of 330 Pa exists due to local atmospheric changes. Calculate the output due to noise and the error in output produced by this noise.

Pressure [kPa]	100	120	140	160	180	200	220	240	260	280	300	320	340	360	380	400
Voltage [V]	1.15	1.38	1.6	1.86	2.1	2.35	2.6	2.89	3.08	3.32	3.59	3.82	4.05	4.29	4.54	4.78

To calculate the output due to noise we must first find the sensitivity of the sensor. Since the output is experimental, we first need to pass a best fit curve through the data. This is necessary since the sensor output is not perfectly linear, as can be seen by plotting the data (**Figure 2.14**). The linear least squares process in **Appendix A** (see **Equation (A.15)**) gives

$$V = aP + b = 0.0122P - 0.0783\text{ V},$$

where P is pressure in kilopascals. Thus the sensitivity is

$$s = \frac{dV}{dP} = a = 0.0122\text{ V/kPa}.$$

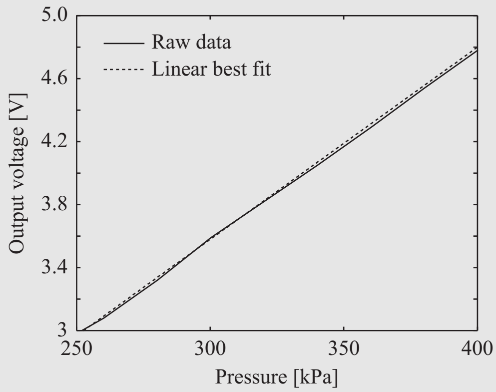

FIGURE 2.14 ■ Plot of sensor output as a function of pressure: raw data and linear best fit.

The output of the sensor at any pressure is that due to pressure and that due to noise (see **Equation (2.11)** or **(2.12)**). Because the curve is linear, we take a convenient point, say, $P = 200$ kPa. The total pressure is 200.33 kPa and the output is

$$V = 0.0122 \times 200.33 - 0.0783 = 2.365726 \text{ V}.$$

Without noise the output is

$$V = 0.0122 \times 200 - 0.0783 = 2.3617 \text{ V}.$$

The output noise is 0.004 V and the error is

$$error = \frac{2.365726 - 2.3617}{2.3617} \times 100 = 0.17\%.$$

EXAMPLE 2.14 Nondestructive testing of materials using a differential inductive probe

One of the reasons to use differential sensors is the cancellation of common-mode effects, including those due to temperature variations and noise, as well as many others. Differential sensors also remove the mean value of the output, leaving only the changes in the output. This is very convenient for further processing of the signal, including amplification, and in some cases it is critical to the operation of the sensor.

As an example, consider a differential eddy current probe used for testing of metals for defects (such as cracks in tubes or in flat surfaces such as the skin of an airplane or a component in an engine). A probe is made of two coils (the sensors), each 1 mm in diameter and separated 2.5 mm apart. Each coil has an inductance L that depends on what is present in the vicinity of the coil. The probe is used to sense a defect in the surface of a steel item by moving it across the material to detect any flaw that may exist. As the leading coil approaches the defect, its inductance goes down, whereas the trailing coil has a higher inductance. In turn, as the leading coil moves past the defect, its inductance increases back to the original value, whereas the trailing coil's inductance goes down. The difference in inductance is the differential output of the probe. **Figure 2.15a** shows the measured inductance of the two coils as they move across the flaw. Note that the behavior of the coils is identical, but the changes occur 2.5 mm apart, as expected.

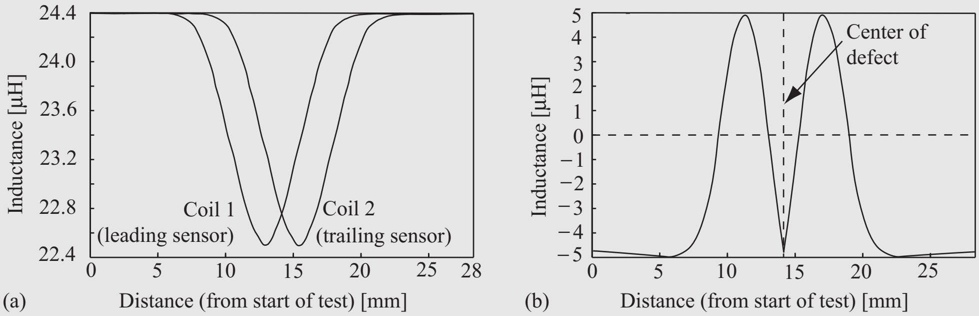

FIGURE 2.15 ■ Eddy current testing of flaws in steel. (a) The inductances of two sensors separated 2.5 mm apart. (b) Differential inductance of the two sensors due to a small flaw.

Taking the differences between the inductances of the two coils at each position as the probe moves leads to the result in **Figure 2.15b**. This differential inductance shows two important aspects. First, the inductance changes around zero—the large inductance (about 24.4 μH) has been removed and all we are left with is the change in inductance due to the defect. This is necessarily so since the inductance due to the solid metal is common to both sensors. Second, the signal is zero when the probe is centered above the flaw—that is, at that point both coils see identical conditions, and since these conditions are common to both, they cause identical changes in the inductance of the coils, cancelling them. This allows identification of the exact position of the defect, an important aspect of testing since the flaw may not be visible with the naked eye or it may be under the surface or under a coating.

It should be noted that the sensitivity has not changed—the change in output per change in input stays the same for each of the two sensors.

There are other ways of connecting sensors. For example, a thermopile ("a pile of thermocouples") is a sensor made of n thermocouples electrically connected in series to increase the electrical output, whereas the input (temperature) to all thermocouples is the same (they are said to be connected in parallel thermally), as shown in **Figure 2.16**. The output now becomes

$$y = y_1 + y_2 + y_3 + \cdots + y_n = (s_1 + s_2 + s_3 + \cdots + s_n) = nsx, \tag{2.17}$$

where it was assumed all thermocouples have identical transfer functions (and therefore sensitivities). The overall sensitivity is

$$S = ns, \tag{2.18}$$

and hence the value of a thermopile.

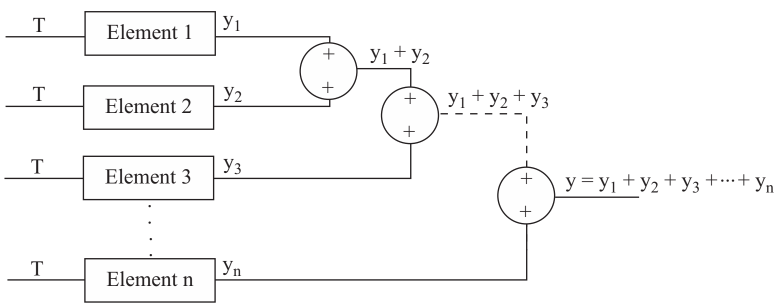

FIGURE 2.16 ■ A sensing system made of n thermocouples connected in series electrically. The output is the sum of the outputs of the individual sensors.

Noise and errors are additive since the outputs are in series. Because the inputs are in parallel thermally, and because the thermocouples are essentially identical, the errors are also identical (or nearly so) and therefore the output error is n times the error (or noise) of a single thermocouple.

2.2.6 Hysteresis, Nonlinearity, and Saturation

Hysteresis (literally lag) is the deviation of the sensor's output at any given point when approached from two different directions (see **Figure 2.17**). Specifically, this means that the output at a given value of stimulus when it increases and when it decreases is different. For example, if temperature is measured, at a rated temperature of 50°C, the output might be 4.95 V when the temperature increases, but 5.05 V when the temperature decreases. This is an error of ±0.5% (for an OFS of 10 V in this idealized example). The sources of hysteresis are either mechanical (friction, slack in moving members), electrical (such as due to magnetic hysteresis in ferromagnetic materials), or due to circuit elements with inherent hysteresis. Hysteresis is also present in actuators and, in the case of motion, is more common than in sensors. There it may manifest itself as positioning errors. Also, hysteresis may be introduced artificially for specific purposes.

Nonlinearity may be either a property of a sensor (see, e.g., **Figure 2.1**) or an error due to deviation of a device's ideal, linear transfer function. A nonlinear transfer function is a property of the device and, as such, is neither good nor bad. One simply has to design with it or around it. However, a nonlinearity error is a quantity that influences the accuracy of the device. It must be known to the designer, must be taken into account, and possibly minimized. If the transfer function is nonlinear, the maximum deviation from linearity across the span is stated as the nonlinearity of the device. However, this measure of linearity is not always possible or desirable. Therefore there are various valid ways of defining the nonlinearity of a sensor or actuator. If the transfer function is close to linear, an approximate line may be drawn and used as the reference linear function. Sometimes this is done simply by connecting the end points (range values) of the transfer function (line 1 in **Figure 2.18**). Another method is to draw a least squares line through the actual curve (line 2 in **Figure 2.18**), usually by first selecting a reasonable number of points on the curve and then, given the selected (or measured) pairs of input and output values (x_i, y_i), draw the line $y = a + bx$ by calculating the slope a and the

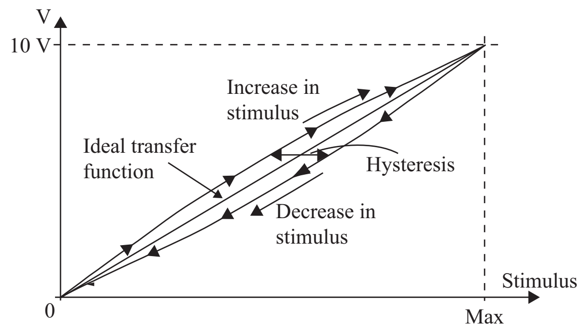

FIGURE 2.17 ■ Hysteresis in a sensor.

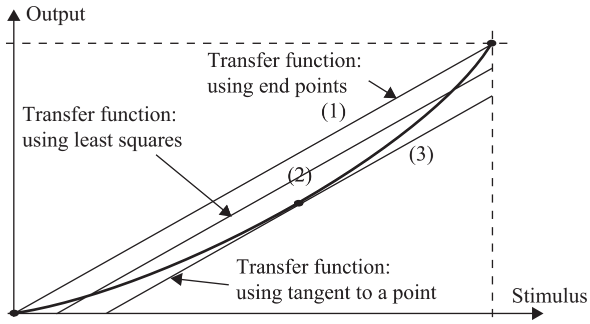

FIGURE 2.18 ■ Linear approximations of nonlinear transfer functions.

axis intercept b as shown in detail in **Appendix A**. The linear best fit to the data (see **Equation (A.15)**) is

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \left\{\sum_{i=1}^{n} x_i\right\}\left\{\sum_{i=1}^{n} y_i\right\}}{n\sum_{i=1}^{n} x_i^2 - \left\{\sum_{i=1}^{n} x_i\right\}^2}, \quad b = \frac{\left\{\sum_{i=1}^{n} y_i\right\}\left\{\sum_{i=1}^{n} x_i^2\right\} - \left\{\sum_{i=1}^{n} x_i\right\}\left\{\sum_{i=1}^{n} x_i y_i\right\}}{n\sum_{i=1}^{n} x_i^2 - \left\{\sum_{i=1}^{n} x_i\right\}^2}. \quad (2.19)$$

This provides a "best" fit (in the least squares sense) through the points of the transfer function and can be used for purposes of measuring the nonlinearity of the actual sensor or actuator. If this method is used, nonlinearity is the maximum deviation from this line.

There are many variations on both of these methods. In some cases a sensor is only expected to operate in a small section of its span. In this case, either method may be applied for that portion of the span. Another method that is sometimes employed is to take a midpoint in this reduced range and draw a tangent to the transfer function through the selected point and use this tangent as the "linear" transfer function for purposes of defining nonlinearity (line 3 in **Figure 2.18**). Needless to say, each method results in different values for nonlinearity and, while these are valid, it is important for the user to know the exact method used.

It should also be noted that in spite of the foregoing, nonlinearity is not necessarily a "bad" thing or something that needs correction. In fact, there are instances in which a nonlinear response is superior to a linear response and there are sensors and actuators that are intentionally and carefully designed to be nonlinear. An example is the common potentiometer used as a volume control, especially in audio systems. Notwithstanding the fact that most current volume control systems tend to be digital and many are linear, our hearing is not linear—in fact, it is logarithmic. This allows the ear to respond to minute pressure changes (as low as 10^{-5} Pa) as well as high pressures (high power sound)—as high as 60 Pa. Normally the range is given between 0 and 130 dB. To accommodate this natural response, potentiometers for volume control were also designed as logarithmic to conform with our ears' response. Even some digital potentiometers are logarithmic. The following example discusses these issues a bit more, but the important point is that this nonlinear response has been designed on purpose to fit a particular need and in this case a nonlinear response is superior.

EXAMPLE 2.15 Rotary logarithmic potentiometer

A 100 kΩ rotary logarithmic potentiometer turns from zero resistance to 100 kΩ in 300° of the slider position. **Figure 2.19a** shows schematically how the potentiometer functions and **Figure 2.19b** shows the resistance as a function of the angular position of the slider relative to the starting point. The resistance at any position of the moving, center tap of the potentiometer (sometimes called a wiper) is calculated as follows:

$$R = 100{,}000\left[1 - \frac{1}{K}\log_{10}\left(\frac{\alpha_{\max} - \alpha + \alpha_{\min}}{\alpha_{\min}}\right)\right] \quad [\Omega],$$

where the normalization factor K is

$$K = \log_{10}\left(\frac{\alpha_{\max} + \alpha_{\min}}{\alpha_{\min}}\right)$$

and $\alpha_{\max} = 300°$ and $\alpha_{\min} = 10°$ represent the maximum and minimum slider angular positions between which a resistance is measurable and α is the slider angular position. This formula ensures the resistance is zero at $\alpha = 0$ and 100 kΩ at $\alpha = 300°$. Note that at, for example, $\alpha = 150°$ the resistance is only 19.75 kΩ, whereas at $\alpha = 225°$ it rises to 38 kΩ.

Notes:

1. There are also antilogarithmic potentiometers in which the resistance increases quickly initially and then levels off.
2. Some logarithmic potentiometers are in fact exponential rather than logarithmic. They are called logarithmic because their response is linear on a log scale.

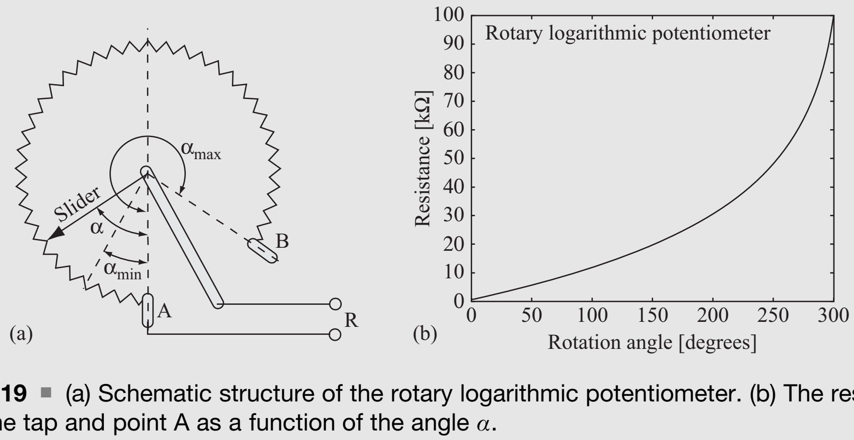

FIGURE 2.19 ■ (a) Schematic structure of the rotary logarithmic potentiometer. (b) The resistance between the tap and point A as a function of the angle α.

Finally, we mention that sometimes one has a choice as to what quantity to use, and a proper choice can make all the difference. In **Chapter 5** we will talk about resistive force sensors, that is, sensors whose resistance changes with the force applied. Naturally in a sensor of this type the resistance is measured, but it so happens that the resistance is highly nonlinear with respect to the applied force. If instead of resistance one measures the conductance (the reciprocal of resistance), the transfer function is perfectly linear. Although the difference between the two representations of the measurand (force) may be dramatic, usage is simple in both cases. If we apply a current source and measure the voltage across the sensor, we obtain a nonlinear response to force. If, on the other hand, we apply a constant voltage source and measure the current through the sensor, the curve is linear and the response of the sensor to force is linear. These issues are shown in **Example 2.16** for an experimentally evaluated resistive force sensor. The choice of selecting the response is not always available, but when it is, it can simplify interfacing considerably.

EXAMPLE 2.16 Linear and nonlinear transfer functions in the same sensor

The response of a resistive force sensor is found experimentally by measuring its resistance as a function of applied force as follows:

Force [N]	0	44.5	89	133	178	222	267	311	356	400	445	489	534
Resistance [Ω]	910	397	254	187	148	122	108	91	80	72	65	60	55

The plot of resistance as a function of force is shown in **Figure 2.20a** and is, as expected, highly nonlinear. The reciprocal plot, that of conductance as a function of force, is shown in **Figure 2.22b** and, within the limitations of measurements, is linear. If a linear response is deemed desirable, one should simply apply a constant voltage source across the sensor and measure its current directly rather than measuring resistance or voltage.

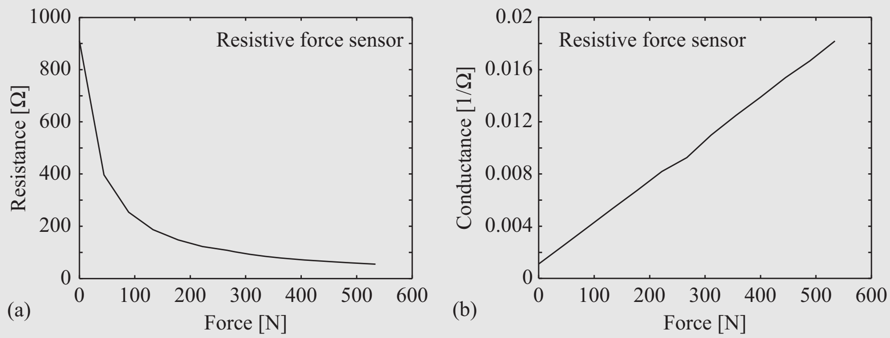

FIGURE 2.20 ■ The transfer function of a resistive force sensor. (a) Plot of the sensor's resistance versus force. (b) Plot of the sensor's conductance versus force.

Saturation refers to the behavior of sensors or actuators when they no longer respond to the input or, more likely, their response is reduced. This usually occurs at or near the ends of their span and indicates that the output is no longer a function of the input or, more likely, is a highly reduced function of the input. In **Figure 2.1** the sensor exhibits saturation at points below T_1 and above T_2, as seen by the flattening of the curves. In both sensors and actuators, one should avoid saturation for two reasons. First, sensing is inaccurate at best and the sensitivity, and often the response, is reduced. Second, saturation may, in some cases, damage the device. In particular, in actuators this may mean that any additional power supplied does not produce an increase in the output power of the device (i.e., sensitivity is reduced), leading to internal heating and possible damage.

2.2.7 Frequency Response, Response Time, and Bandwidth

Frequency response (also called the **frequency transfer function**) of a device indicates the ability of the device to respond to a harmonic (sinusoidal) input. Typically the frequency response shows the output (as a magnitude of the output or gain) of a device as a function of the frequency at the input, as shown in **Figure 2.21**. Sometimes the phase of the output is also given (the pair amplitude gain-phase responses is known as the Bode diagram). The frequency response is important in that it indicates the range of frequencies of the stimulus for which the output is adequate (i.e., does not deteriorate or increase the error due to the inability of the device to operate at a frequency or range of frequencies). For sensors and actuators that are required to operate over a range of frequencies, the frequency response provides three important design parameters. One is the **bandwidth** of the device. This is the frequency range between the two pre-agreed-upon points A and B in **Figure 2.21**. These points are almost always taken as the half-power points (at which power is half that of the flat region). The gain (magnitude) at the half-power points is $1/\sqrt{2}$ of the gain in the flat region, or 70.7%. Often the frequency response is given in decibels, in which case the half-power points are points at which the gain is 3 dB down ($10 \times \log 0.5 = -3$ dB or $20 \times \log(\sqrt{(2)}/2) = -3$ dB). The second parameter that is sometimes used is the **useful frequency range** or **flat frequency range** (or **static range**), which, as its name implies, is that portion of the bandwidth that is flat. However, most devices do not have a frequency response anywhere near the ideal response shown in **Figure 2.21**. Therefore the useful frequency range may be

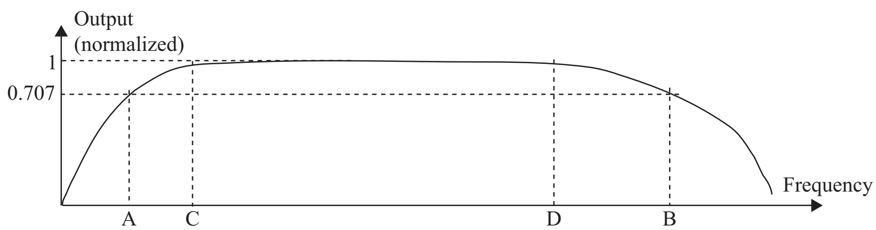

FIGURE 2.21 ■ Frequency response of a device showing the half-power points.

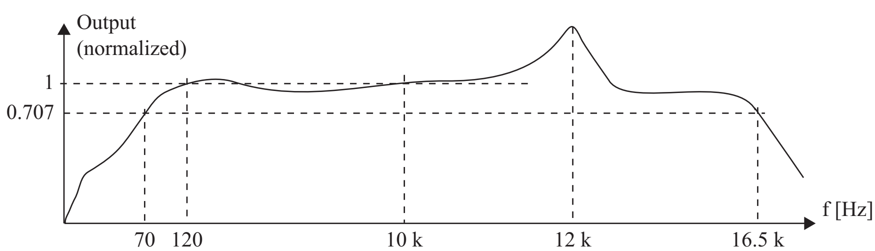

FIGURE 2.22 ■ Definition of bandwidth for a device with nonflat frequency response (frequency axis not to scale).

considerably smaller than the bandwidth (depending on how flat it needs to be for the application) or it may be a compromise between flatness and width. An example is shown in **Figure 2.22**, where the frequency response of a hypothetical loudspeaker is shown. The bandwidth extends between 70 Hz and 16.5 kHz and most often is indicated as 16.5 kHz$-$70 Hz $=$ 16,430 Hz (the difference between the half-power points). The response at 12 kHz is said to be resonant (maximum in this case, but can also be minimum). The flat region is not entirely flat, but we could reasonably take the range between 120 Hz and 10 kHz as "flat." In this region the speaker will reproduce the input most faithfully. The half-power points on the frequency response curve are viewed as **cutoff frequencies**, essentially indicating that beyond these points the device is "useless." Of course, these are arbitrary points to a large extent and the device can operate beyond these points, but with a reduced response. In some cases the lower cutoff frequency does not exist, indicating that the device responds down to DC.

Related to frequency response is the **response time** (or delay time) of the device, which indicates the time needed for the output to reach steady state (or a given percentage of steady state) for a step change in input. This is often specified for slow-responding devices such as temperature sensors or thermal actuators. Typically the response time will be given as the time needed to reach 90% of steady-state output upon exposure to a unit step change in input. The response time of the device is due to the inertia of the device (mechanical, thermal, electrical). For example, in a temperature sensor, it may simply be due to the time needed for the sensor's body to reach the temperature it is trying to measure (thermal time constant) or may be due to the electrical time constants inherent in the device due to capacitances and inductances, or, most likely, due to both. Clearly this means that the higher the response time, the less the sensor can follow rapid changes in the stimulus, and consequently the narrower its frequency response (bandwidth). Response time is an important design parameter that the engineer must take into account. As a rule, since response time is mostly related to mechanical time constants, and these are in general related to physical dimensions, smaller sensors tend to have shorter response times, whereas bulky sensors tend to respond slower (longer response times). Response time is most often specified with devices that respond slowly. Fast-acting devices will be specified in terms of frequency response.

EXAMPLE 2.17 Frequency response of a magnetic sensor

The frequency response of a magnetic sensor used to detect flaws in conducting structures is shown in **Figure 2.23**. This sensor is called an eddy current sensor (really a rather simple coil) and is a common sensor in testing of tubes for internal or external flaws. The frequency response is rather narrow, indicating that the sensor is resonant, in this case with a center frequency of about 290 kHz. Nevertheless, the resonance is not very sharp, indicating a lossy resonant circuit. In a device of this type one tries to operate at a frequency around the resonant frequency. The output of the sensor is typically a voltage when fed with a constant current source or a current when fed with a constant voltage source (i.e., either the voltage across the sensor or the current in the sensor is measured, depending on the way the sensor is fed). The amplitude and phase of the output are then related to the size, type, and location of flaws. (Eddy current sensors will be discussed in **Chapter 5**).

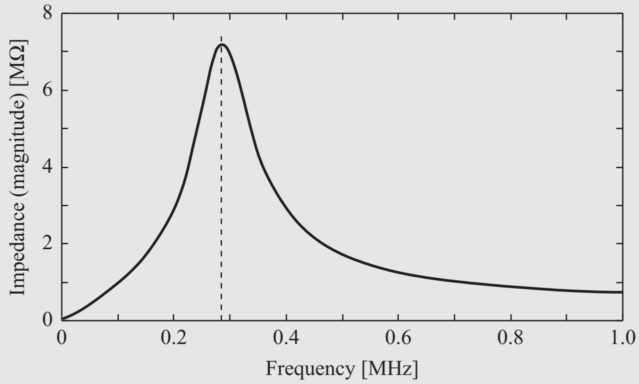

FIGURE 2.23 ■ Eddy current sensor frequency response.

2.2.8 Calibration

Calibration is the experimental determination of the transfer function of a sensor or actuator. Typically calibration of a device will be needed when the transfer function is not known or, more likely, when the device must be operated at tolerances below those specified by the manufacturer. Since tolerances indicate the maximum (sometimes typical) deviations of the device's transfer function from the ideal, if the device needs to operate at lower tolerances, we must specify the exact transfer function for the specific device. For example, we may wish to use a thermistor with a 5% tolerance on a full scale from 0°C to 100°C (i.e., the variations in measurements are ±5°C) to measure temperature with an accuracy of, say, ±0.5°C. The only way this can be done is by first establishing the transfer function of the sensor. For best results, this must be done for each device. There are two ways this can be accomplished.

One method assumes that the equation of the transfer function is known, in which case the constants in the equation must be determined experimentally. Suppose that the thermistor above has a linear transfer function between the range points given as $R = aT + b$, where T is the measured temperature and R is the resistance of the sensor, with a and b constants. To establish the transfer function for the thermistor we must specify the constants a and b, and for this we require two measurements at two different temperatures, T_1 and T_2. Then we can write

$$V_1 = aT_1 + b, \quad V_2 = aT_2 + b. \tag{2.20}$$

These are solved for a and b as

$$a = \frac{V_1 - V_2}{T_1 - T_2}, \quad b = V_1 - \frac{V_1 - V_2}{T_1 - T_2} T_1 = V_1 - aT_1. \tag{2.21}$$

This relation can now be supplied to the processor and from it the measured temperature is deduced through the transfer function $R = aT + b$.

If the transfer function is described by a more complex function, more measurements may be needed, but in all cases the constants in the relation must be determined.

For example, suppose that an actuator's output force is given as $F = aV + bV^2 + cV^3 + d$. The constants a, b, c, and d need to be determined and this will require four measurements and solution of a system of four equations.

In the calibration process one should be careful and choose measurement points within the span of the device and, especially for nonlinear transfer functions, space the measurement points more or less equally within the span. For linear functions, any two points will do, but even here they should not be too close to each other.

The second method is to assume no knowledge of the transfer function and determine it by a series of experiments. Typically a number of measurements will be needed measuring T_i and reading the resulting resistance R_i. These represent points of the transfer function. The points are plotted and a curve (best fit curve) passed through the points. It may turn out that the points define a (more or less) linear curve, in which case the linear least squares method described in **Appendix A** (**Equation (A.15)**) may be used. Otherwise a polynomial fit through the points may be needed (see **Equation (A.24)** in **Appendix A**). Alternatively, the points may be supplied to the processor (especially if this is a microprocessor) as a table and the processor can then be programmed to retrieve these values and perhaps to interpolate between them for stimuli that fall between two measurements. This linear piecewise approximation may be quite sufficient, especially if the calibration uses a reasonably large number of points.

Calibration is a critical step in the use of sensors or actuators and should be undertaken with the utmost care. Measurements must be meticulous, instruments as accurate as possible, and conditions as close as possible to those under which the sensor or actuator will operate. One should also establish the errors in calibration or, at the very least, have a good estimate of the errors.

2.2.9 Excitation

Excitation refers to the electrical supply required for operation of a sensor or actuator. It may specify the range of voltages under which the device should operate (say, 2–12 V), the range of current, power dissipation, maximum excitation as a function of temperature, and sometimes frequency. Together with other specifications (such as mechanical properties and electromagnetic compatibility [EMC] limits), it defines the normal operating conditions of the sensor. Failure to follow rated values may result in erroneous outputs or premature failure of the device.

2.2.10 Deadband

Deadband is the lack of response or insensitivity of a device over a specific range of the input. In this range, which may be small, the output remains constant. A device should not operate in this range unless this insensitivity is acceptable. For example, an actuator that is not responding to inputs in a small range around zero may be acceptable, but one that "freezes" over the normal range may not be.

2.2.11 Reliability

Reliability is a statistical measure of the quality of a device that indicates the ability of the device to perform its stated function, under normal operating conditions, without failure, for a stated period of time or number of cycles. Reliability may be specified in hours or years of operation or as number of cycles or number of failures in a sample. Electronic components including sensors and actuators are rated in a number of ways.

The **failure rate** is the number of components that fail per given time period, typically per hour. A more common method of specifying the reliability of devices is in **mean time between failures** (MTBF). MTBF is the reciprocal of the failure rate: MTBF = 1/(failure rate). Another common term used to specify reliability is the **failure in time** (FIT) value. This measure gives the number of failures in 10^9 device-hours of operation. The device-hour figure can be made of any number of devices and hours as long as the product is 10^9 (say, 10^6 devices tested for 1000 hours). It can also be done with fewer device-hours and scaled to the required value. For example, one may test 1000 devices for 1000 hours and multiply the result by 1000.

Reliability data are usually provided by manufacturers and are based on accelerated lifetime testing conducted by the manufacturer. Although specification sheets do not usually provide much data about reliability or the methods used to obtain reliability data, most manufacturers will supply these data upon request and some may also have certification data based on standards of testing, when applicable.

It should be noted that reliability is heavily influenced by the operating conditions of devices. Elevated temperatures, higher voltages and currents, as well as environmental conditions (such as humidity) reduce reliability, sometimes dramatically. Any condition exceeding rated values must be taken into account and the reliability data derated accordingly. In some cases these data are available from manufacturers or professional organizations dedicated to issues of reliability. Calculators are available allowing the user to estimate reliability.

EXAMPLE 2.18 Failure rate

To test a component, 1000 identical components are tested for 750 hours and 8 of them fail during the test. The failure rate is

$$FR = \frac{8}{750 \times 1{,}000} = 1.067 \times 10^{-5}.$$

That is, the failure rate is 1.067×10^{-5} failures/h. The mean time between failures is MTBF = 93,750 hours. One can also estimate the FIT rate. Since the device-hour figure is $750 \times 1000 = 750{,}000$ device-hours, the FIT rate (number of failures for 10^9 device-hours) is

$$FIT = \frac{8}{750 \times 1{,}000} \times 10^9 = 10{,}666.$$

This (fictitious) component has extremely low reliability. Typical values for FIT rates are 2–5 and MTBF is usually measured in billions of hours.

2.3 | PROBLEMS

The transfer function

2.1 **Error in the simplified transfer function.** In **Example 2.1**, suppose that the transfer function is simplified to a third-order function by neglecting all terms except for the first three. Calculate the largest error expected over the range between 0°C and 1800°C.

2.2 **Transfer function of a position sensor.** The transfer function of a small position sensor is evaluated experimentally. The sensor is made of a very small magnet and the position with respect to the centerline (see **Figure 2.24**) is sensed by the horizontal, restoring force on the magnet. The magnet is held at a fixed distance, h, from the iron plate. The measurements are given in the table below.

a. Find the linear transfer function that best fits these data.

b. Find a transfer function in the form of a second-order polynomial ($y = a + bf + cf^2$), where y is the displacement and f is the restoring force by evaluating the constants a, b, and c.

c. Plot the original data together with the transfer functions in (a) and (b) and discuss the errors in the choice of approximation.

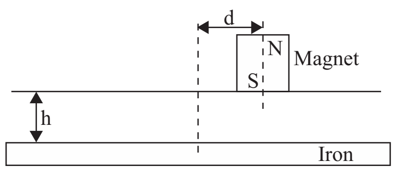

FIGURE 2.24 ■ A simple position sensor.

Displacement, d [mm]	0	0.08	0.16	0.24	0.32	0.4	0.48	0.52
Force [mN]	0	0.578	1.147	1.677	2.187	2.648	3.089	3.295

2.3 **Analytic form of transfer function.** In certain cases the transfer function is available as an analytic expression. One common transfer function used for resistance temperature sensors (to be discussed in **Chapter 3**) is the Callendar–Van Duzen equation. It gives the resistance of the sensor at a temperature T as

$$R(T) = R_0\left(1 + AT + BT^2 - 100CT^3 + CT^4\right) \quad [\Omega],$$

where the constants A, B, and C are determined by direct measurement of resistance for the specific material used in the sensor and R_0 is the temperature of the sensor at 0°C. Typical temperatures used for calibration are the oxygen point (−182.962°C; the equilibrium between liquid oxygen and its vapor), the triple point of water (0.01°C; the point of equilibrium temperature between ice, liquid water, and water vapor), the steam point (100°C; the equilibrium point between water and vapor), the zinc point (419.58°C; the equilibrium point between solid and liquid zinc), the silver point (961.93°C), and the gold point (1064.43°C), as well as others. Consider a platinum resistance sensor with a nominal resistance of 25 Ω at 0°C. To calibrate the sensor its resistance is measured at the oxygen point as 6.2 Ω, at the steam point as 35.6 Ω, and at the zinc point as 66.1 Ω. Calculate the coefficients A, B, and C and plot the transfer function between −200°C and 600°C.

Impedance matching

2.4 Loading effects on actuators. Suppose an 8 Ω loudspeaker is driven from an amplifier under matched conditions, meaning that the amplifier's output impedance equals 8 Ω. Under these conditions the amplifier transfers maximum power to the loudspeaker. The user decides to connect a second, identical speaker in parallel to the first to better distribute sound in the room. The amplifier supplies an output voltage $V = 48$ V.

a. Show that the total power on the two speakers is lower than that supplied to a single speaker.

b. If one desires to maintain the same power, what must be the impedance of the speakers assuming they are still connected in parallel to the amplifier?

2.5 Loading effects. The ratings of a piezoelectric sensor indicate that at a certain stimulus the sensor provides an output of 150 V at no load. The short-circuit current of the sensor is measured and found to be 10 μA (this is done by shorting the output through a current meter). To measure the output voltage of the sensor, a voltmeter with an internal impedance of 10 MΩ is connected across the sensor.

a. Calculate the actual reading of the voltmeter and the error in reading the stimulus due to the impedance of the voltmeter.

b. What must be the impedance of the voltmeter to reduce the error in reading below 1%?

2.6 Impedance matching effects. A pressure sensor produces an output varying between 0.1 and 0.5 V as the pressure sensed varies from 100 to 500 kPa. To read the pressure the sensor is connected to an amplifier with an amplification of 10 so that the output varies between 1 and 5 V for ease of interpretation of the sensed pressure. If the sensor has an internal impedance of 1 kΩ and the input impedance of the amplifier is 100 kΩ, what is the voltage range at the output of the amplifier?

2.7 Power output of an electrical motor. The torque of a DC motor is linear with the speed of the motor and given in **Figure 2.25**. Find the power transfer function of

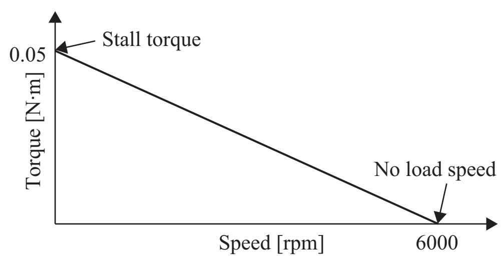

FIGURE 2.25 ■ Speed–torque curve of a DC motor.

the motor, that is, the relation between speed and power. Show that maximum mechanical power is obtained at half the no-load speed and/or half the stall torque.

Note: Power is the product of torque and angular velocity.

2.8 Power transfer to an actuator and matching. A power amplifier may be modeled as an ideal voltage source with a series internal impedance whereas a load is modeled as an impedance, both shown in **Figure 2.26**. The internal impedance of the amplifier is $Z_{in} = 8 + j2\ \Omega$ at frequency f.

a. Calculate and plot the power supplied to the load if $V_0 = 48$ V for a resistive load varying between 0 and 20 Ω.

b. Calculate and plot the power supplied to the load if $V_0 = 48$ V for a load varying between $Z_L = 8 + j0\ \Omega$ to $Z_L = 8 + j20\ \Omega$.

c. Show that the maximum power to the load is transferred if $Z_L = Z_{in}{}^* = 8 - j2\ \Omega$.

d. How do you explain the result in (c) from a physical point of view?

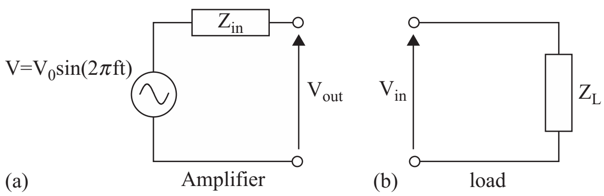

FIGURE 2.26 ■ Power transfer and matching. (a) Model of the amplifier output. (b) The load.

2.9 Matching of frequency-dependent impedances. An actuator is given with a resistance of 8 Ω at 100 Hz. Its inductance is measured and found to be 1 mH. The actuator is used at frequencies ranging from 10 to 2000 Hz. In attempting to match the actuator to a driving source the following are proposed:

1. Use a driver with an output resistance equal to 8 Ω and a source with peak voltage of 12 V.
2. Use a driver with an output impedance of $8 + j0.006f\ \Omega$, where f is the frequency [Hz] and a source with peak voltage of 12 V.
3. Use a driver with an output impedance of $8 - j0.006f\ \Omega$, where f is the frequency [Hz] and a source with peak voltage of 12 V.

a. Calculate and plot the output power supplied to the actuator for the three proposed drivers. Which one is better and why?

b. What are the requirements of an ideal driver for the given load?

Range, span, input and output full scale, and dynamic range

2.10 The human ear is a uniquely sensitive sensor. Its range is given either in pressure or in power per unit area. The ear can sense pressures as low as 2×10^{-5} Pa (on the order of a billionth of the atmospheric pressure) and can still function properly at levels of 20 Pa at the high end (the threshold of pain). Alternatively, the range can be specified from 10^{-12} W/m^2 to 10 W/m^2. Calculate the dynamic range for pressure and power.

2.11 **The human eye.** The human eye is sensitive from roughly 10^{-6} cd/m^2 (dark night, rod-dominated vision, essentially monochromatic) to about 10^6 cd/m^2 (bright sunlight). Calculate the dynamic range of the eye.

2.12 A loudspeaker is rated at 10 W, that is, it can produce 10 W of acoustic power. Since it is an analog actuator, the minimum range point is not well defined but there is a minimum power necessary to overcome friction. We will assume here that it is 10 mW. What is the dynamic range of the loudspeaker?

Note: A loudspeaker's dynamic range is measured in various ways, some of them intended to boost marketability rather than to describe the physical properties of the loudspeaker.

2.13 A digital frequency meter needs to measure the frequency of a microwave sensor whose frequency varies between 10 MHz and 10 GHz in increments of 100 Hz. What is the dynamic range of the frequency meter?

2.14 A liquid crystal display is said to have a contrast ratio of 3000:1, giving the ratio between the highest and lowest luminance it can display. Luminance is a measure of power per unit area per solid angle. Calculate the dynamic range of the display in decibels.

Note: In displays, the dynamic range is usually given as the contrast ratio, whereas in digital cameras it is given as *f*-stops, but these can always be written in decibels if necessary. For example, a digital camera with a contrast of 1024:1 is said to have a dynamic range of 2^{10} or 10 *f*-stops.

2.15 Digital signal processors, especially those handling audio and video data, must have large dynamic ranges. If the dynamic range of the A/D conversion of a signal processor must be at least 89 dB, what is the number of bits required of the D/A converter?

Sensitivity, accuracy, errors, and repeatability

2.16 **Linear approximation of nonlinear transfer function.** The response of a temperature sensor is given as

$$R(T) = R_0 e^{\beta\left(\frac{1}{T}-\frac{1}{T_0}\right)} \quad [\Omega],$$

where R_0 is the resistance of the sensor at temperature T_0 and β is a constant that depends on the material of the sensor. $T_0 = 20°\text{C}$. Temperatures T and T_0 are in K. Given: $R(T) = 1000\ \Omega$ at 25°C and 3000 Ω at 0°C. The sensor is intended for use between −45°C and 120°C.

a. Evaluate β for this sensor and plot the sensor transfer function for the intended span.
b. Approximate the transfer function as a straight line connecting the end points and calculate the maximum error expected as a percentage of full scale.
c. Approximate the transfer function as a linear least squares approximation and calculate the maximum error expected as a percentage of full scale.

Hysteresis

2.17 **Hysteresis in a torque sensor.** A torque sensor is calibrated by applying static torque to it (i.e., a certain torque is applied, the sensor response is measured, and then the torque is increased or decreased to measure another point on the curve). The following data are obtained. The first set is obtained by increasing torque, the second by decreasing it.

Applied torque [N·m]	2.3	3.14	4	4.84	5.69	6.54	7.39	8.25	9.09	9.52	10.37	10.79
Sensed torque [N·m]	2.51	2.99	3.54	4.12	4.71	5.29	5.87	6.4	6.89	7.1	7.49	7.62

Applied torque [N·m]	10.79	10.37	9.52	9.09	8.25	7.39	6.54	5.69	4.84	4	3.14	2.3
Sensed torque [N·m]	7.68	7.54	7.22	7.05	6.68	6.26	5.8	5.29	4.71	4.09	3.37	2.54

a. Plot the transfer function of the torque sensor using a second-order least squares approximation.
b. Calculate the maximum error due to hysteresis as a percentage of full scale.

2.18 **The Schmitt trigger.** Hysteresis is not necessarily a negative property. The use of hysteresis in electronic circuits can accomplish specific purposes. One of these is the Schmitt trigger. It is an electronic circuit whose output voltage (V_{out}) changes according to the input voltage (V_{in}) as follows:

$$\textit{if } V_{in} \geq 0.5\,V_0,\ V_{out} = 0$$

$$\textit{if } V_{in} \leq 0.45\,V_0,\ V_{out} = V$$

a. Draw the transfer function of the device for $0 \leq V_{in} \leq V_0$ with $V_0 = 5$ V.
b. Suppose the input is a sinusoidal voltage $V_{in} = 5\sin 2\pi f$, where t is time and $f = 1000$ Hz is the frequency of the signal. Plot the input and the output as a function of time. What is the purpose of the Schmitt trigger in this case?

2.19 **Mechanical hysteresis.** Springs are often used in sensors, particularly in sensing forces, taking advantage of the fact that under an applied force the spring contracts (or expands) based on the formula $F = -kx$ (Hooke's law), where F is the applied force, x is the spring contraction or expansion, and k is a constant called the spring constant. The negative sign indicates that compression reduces the length of the spring. However, the spring constant is force dependent and has slightly different variations under compression and expansion. Consider the spring in **Figure 2.27a** and its force–displacement curve given in **Figure 2.27b**.

a. Explain the meaning of the curve and, in particular, that of the horizontal lines on which additional force does not change the displacement.
b. What is the maximum error in displacement as a percentage of full scale due to hysteresis?
c. What is the maximum error in force as a percentage of full scale due to hysteresis?

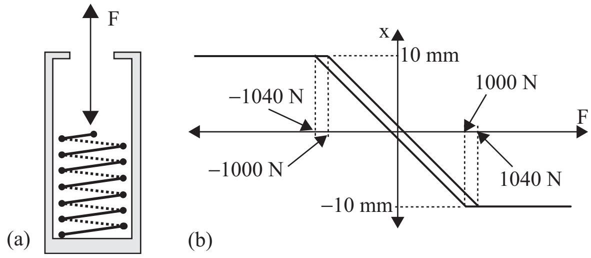

FIGURE 2.27 ■ Mechanical hysteresis in a spring. (a) Spring and applied force. (b) Response.

2.20 **Hysteresis in thermostats.** Hysteresis is often intentionally built into sensors and actuators. A simple example is the common thermostat. These devices are designed to switch off at a certain temperature and then switch back on at a certain temperature. The switch-on and switch-off temperatures must be different, otherwise the thermostat's status is undetermined and it will switch on and off rapidly when the set temperatures are reached. The hysteresis can be mechanical, thermal, or electronically imposed. Consider a thermostat intended to control temperature in a home. The set temperature is 18°C but the thermostat has a ±5% hysteresis.

a. For a temperature between 15°C and 24°C, determine the transfer function of the thermostat if it is used to turn on a heater to keep the room warm.
b. For the same range as in (a), determine the transfer function of the thermostat if it is to turn on an air conditioner to keep the room cool.

In both cases, show when the thermostat switches on and off.

Deadband

2.21 **Deadband due to linkage slack.** A linear rotary potentiometer of nominal value 100 kΩ is used to sample the voltage across a sensor (such as a microphone). The shaft has a rotational slack of 5°, that is, if one rotates it in one direction to a certain point and then rotates it in the opposite direction, the shaft must be rotated 5° before the slider responds. If the full-scale rotation is 310°, calculate the error due to the slack in the linkage in terms of resistance and as a percentage of full scale.

Reliability

2.22 **Reliability.** An electronic component's data sheet shows a mean time between failure (MTBF) of 4.5×10^8 hours when tested at 20°C. At 80°C the MTBF decreases to 62,000 hours. Calculate the failure rate (FR) and the failure in time (FIT) value for the component at the two temperatures.

2.23 **Reliability.** In testing a pressure sensor for reliability, 1000 sensors are tested for 850 hours. If six sensors fail during that period, what is the MTBF of the sensor?